

# Houda Bouamor, PhD

*Assistant Professor of Information Systems*

Work address: Carnegie Mellon University in Qatar  
Education City, P.O Box 24866, Doha, Qatar  
Home address: 122, Rue de Vaugirard, 75006. Paris  
Phone: (+974) 33.15.98.56  
Email: hbouamor@qatar.cmu.edu

## Profile

---

- PhD. in Computer Science, major in Natural Language Processing (NLP)
- Experienced in:
  - Teaching large-enrollment introductory courses, as well as higher level electives to students majoring on Computer Science, Information Systems, Biological Sciences and Business Administration
  - Supervising student research activities: junior and senior thesis and projects.
  - Natural Language Processing, applied machine learning and deep learning
  - Collecting, building and maintaining corpora to train and validate NLP models
  - Working with unstructured data both formal and informal genres

## Research Interests

- Applied machine learning and Deep learning
- Supervised and unsupervised learning methods
- NLP: monolingual and multilingual processing, natural language processing for morphologically complex languages, dialectal Arabic processing, low-resource languages processing, named entity recognition and translation, statistical and neural machine translation, machine translation evaluation and confidence estimation, paraphrasing, dialect identification, user-generated content processing and social media analysis, sentiment analysis, eye-tracking for NLP, automatic spelling correction, financial data processing, natural language generation.
- Data Science: Data collection, Data processing, Data extraction and visualization, Data analysis, etc. (with a focus on textual data)

## Key Achievements

- Top-tier journal and conference publications (ACM, ACL, EMNLP, EACL, NAACL, COLING)

- Grant of 1 million USD funding for a research project on Dialectal Arabic Processing
- Anita Borg Institute scholarship to attend the Grace Hopper Conference
- Best paper award in the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013)
- Full PhD scholarship from the Ministry of Higher Education and Research in France

## Education

---

2008 – 2012	<b>PhD in Computer Science</b> <i>University Paris-Sud 11, Orsay, France</i> Major in Computational Linguistics
2007 – 2008	<b>Master of Science in Computer Science</b> <i>University Paris-Est Marne-La-Vallée, France</i> Major in Scientific and Technical Information Processing
2003 – 2007	<b>Bachelor of Computer Science</b> <i>School of Business, University of Manouba, Tunis, Tunisia</i> Major in Computer Science Applied to Business Management

## Research and Work Experience

---

May. 2018– Feb. 2019	<b>Chief Research Scientist and NLP Expert</b> Fortia Financial Solutions <i>Financial Data Analysis and Processing, Financial Data Generation</i>
Oct. 2014 – Aug. 2018	<b>Senior Researcher</b> Carnegie Mellon University in Qatar <i>Arabic Natural Language Processing, Colloquial Arabic processing, Statistical Machine Translation, Sentiment Analysis, Automatic Text Correction, Machine Translation Evaluation, Quality Estimation, Arabic diacritization</i>
Sep. 2012 – Oct. 2014	<b>Postdoctoral Research Associate</b> Carnegie Mellon University in Qatar <i>Arabic Natural Language Processing, Colloquial Arabic processing, Statistical Machine Translation, Sentiment Analysis</i>

- 2008 – 2012 **Research Assistant**  
LIMSI-CNRS, Orsay, France  
Information, Language, writtEn and Signed Group (ILES)  
*PhD thesis title:* Technique Combination for Paraphrase Alignment :  
Contribution to Pinpointing Sub-sentential Paraphrases  
*Supervisors:* Aurélien Max and Anne Vilnat
- Apr. 2007 – Sep. 2008 **Research Intern**  
CEA-LIST , France  
Multilingual Multimedia knowledge engineering lab (LIC2M)  
*Master thesis title:* Mining a Multilingual Geographical Gazetteer  
from the Web  
*Supervisors:* Adrian Popescu, Pierre-Alain Moellic
- June – Aug. 2007 **Research Intern**  
LLACAN-CNRS , France  
*Topic:* Arabic Natural Language Processing: Interactive Annotation  
of Arabic Corpus

## Teaching Experience

---

- March 2019 – **Assistant Teaching Professor**  
*Courses:* 67-272 *Application Design and Development,*  
*Introduction to Data Analysis*  
Carnegie Mellon University in Qatar
- August 2015 – August 2018 **Visiting Assistant Professor**  
*Courses:* *Principles of Computing (15-110), Fundamentals of Programming*  
*and Computer Science (15-112), Introduction to Data Structures (15-121) and*  
*Introduction to Machine Learning (10-601)*  
Carnegie Mellon University in Qatar
- Summer 2016 **Visiting Assistant Professor**  
*Course:* *Introduction to Data Structures (15-121)*  
Carnegie Mellon University, Pittsburgh
- Summer 2015 **Instructor**  
*Courses:* *Fundamentals of Programming and Computer Science(15-112)*  
Carnegie Mellon University in Qatar
- Fall 2013 **Guest Lecture**  
*Lecture:* *Introduction to Text Processing: POS Tagging*  
Carnegie Mellon University in Qatar
- 2011 – 2012 **Teaching Assistant (ATER)**  
*Courses:* *Algorithms and, C and Java Programming Languages*  
Polytech' Paris Sud & Paris Sud University, UFR des Sciences

2008 – 2011

**Teaching Assistant (Monitrice)**

Courses: *Algorithms, Information System Design, Databases, Java Programming language*  
 Institut Universitaire de Technologies (IUT) d'Orsay,  
 University Paris-Sud 11

---

**Publications****Patents**

1. **Houda Bouamor**, Sira Ferradans, Guillaume Hubert and Abderrahim Ait-Azzi, "Data-to-text Generation at Fortia". Filed at the US Office. Feb 2019.

The following publications are also available from my Google citation page.

**Refereed Journal Publications**

1. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Multi-technique paraphrase alignment: A contribution to pinpointing sub-sentential paraphrases", *ACM Transactions on Intelligent Systems and Technology (Special Issue on Paraphrasing)*. 2013. Volume 4. Issue 3.
2. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Étude bilingue de l'acquisition et de la validation automatiques de paraphrases sous-phrastiques" *Traitement Automatique des Langues*, 2012, volume 53.1  
*English title: "A bilingual study of sub-sentential paraphrase automatic acquisition and validation."*
3. Camille Dutrey, **Houda Bouamor**, Delphine Bernhard, Aurélien Max, "Local modifications and paraphrases in Wikipedia's revision history. " *SEPLN journal (Revista de Procesamiento del Lenguaje Natural)*, volume 46:51-58.

**Refereed International Conference and Workshop Publications**

1. **Houda Bouamor**, Sabit Hassan and Nizar Habash, "The MADAR Shared Task on Arabic Fine-Grained Dialect Identification". *WANLP 2019. The Fourth Arabic Natural Language Processing Workshop*. Florence, Italy, 2019
2. Nizar Habash, **Houda Bouamor** and Christine Chung, "Gender Aware Arabic Natural Language Processing: First Person Gender Identification and Reinflection". *GenderBiasNLP 2019. The 1st ACL Workshop on Gender Bias for Natural Language Processing*, Florence, Italy: 2019.
3. Alexander Erdmann, Salam Khalifa, Mai Oudah, Nizar Habash and **Houda Bouamor**, "A Little Linguistics Goes a Long Way: Unsupervised Segmentation with Limited Language Specific Guidance". *SIGMORPHON 2019. The 16th ACL SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, Florence, Italy: 2019.
4. Ossama Obeid, Mohammad Salameh, **Houda Bouamor**, and Nizar Habash, "ADIDA: Automatic Dialect Identification for Arabic". *NAACL 2019 (demo Session). The Conference of the North American Association for Computational Linguistics*, Minneapolis, USA: 2019.

5. Abderrahim Ait-Azzi and **Houda Bouamor**, "Fortia1 @ the NTCIR-14 FinNum Task: Enriched Sequence Labeling for Numeral Classification". *NTCIR-14 2019. The 14th NTCIR (2018-2019) Evaluation of Information Access Technologies (To appear)*, Tokyo, Japan: 2019.
6. Mohammad Salameh, **Houda Bouamor** and Nizar Habash, "Fine-Grained Arabic Dialect Identification". *COLING 2018. The 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA: 2018.
7. **Houda Bouamor** and Hassan Sajjad. "H2@BUCC18: Parallel Sentence Extraction from Comparable Corpora Using Multilingual Sentence Embeddings". *BUCC 2018. 11th workshop on Building and Using Comparable Corpora*, Miyazaki, Japan. 2018
8. **Houda Bouamor**, Nizar Habash, Mohammad Salameh, Wajdi Zaghrouani, Owen Rambow, Dana Abdulrahim, Ossama Obeid, Salam Khalifa, Fadhl Eryani, Alexander Erdmann and Kemal Oflazer. "The MADAR Arabic Dialect Corpus and Lexicon". *LREC 2018. 11th edition of the Language Resources and Evaluation Conference*. Miyazaki, Japan: 2018.
9. Ossama Obeid, Salam Khalifa, Nizar Habash, **Houda Bouamor**, Wajdi Zaghrouani and Kemal Oflazer. "MADARi: A Web Interface for Joint Arabic Morphological Annotation and Spelling Correction". *LREC 2018. 11th edition of the Language Resources and Evaluation Conference*. Miyazaki, Japan: 2018.
10. Nizar Habash, Salam Khalifa, Fadhl Eryani, Owen Rambow, Dana Abdulrahim, Alexander Erdmann, Reem Faraj, Wajdi Zaghrouani, **Houda Bouamor**, Nasser Zalmout, Sara Hassan, Faisal Al shargi, Sakhar Alkhereyf, Basma Abdulkareem, Ramy Eskander, Mohammad Salameh and Hind Saddiki. "Unified Guidelines and Resources for Arabic Dialect Orthography". *LREC 2018. 11th edition of the Language Resources and Evaluation Conference*. Miyazaki, Japan: 2018.
11. Alexander Erdmann, Nizar Habash, Dima Taji and **Houda Bouamor**. "Low Resourced Machine Translation via Morpho-syntactic Modeling: The Case of Dialectal Arabic". *MT Summit 2017. MT Summit XVI, vol.1: Research Track*, Nagoya, Japan: 2017.
12. Francisco Guzmàn, **Houda Bouamor**, Ramy Baly and Nizar Habash. "Machine Translation Evaluation for Arabic using Morphologically-enriched Embeddings". *COLING 2016. The 26th International Conference on Computational Linguistics*, Osaka, Japan: 2016.
13. Wajdi Zaghrouani, Nizar Habash, **Houda Bouamor**, Ossama Obeid, Sawsan Alqahtani, Mona Diab and Kemal Oflazer. "Filtering Dialectal Arabic Text in Two Large Scale Annotation Projects". *W-NUT 2016. The 2nd Workshop on Noisy User-generated Text*. Osaka, Japan: 2016
14. Wajdi Zaghrouani, Abdelati Hawwari, Sawsan Alqahtani, **Houda Bouamor**, Mahmoud Ghoneim, Mona Diab and Kemal Oflazer. "Using Ambiguity Detection to Streamline Linguistic Annotation". *CL4LC. Workshop on Computational Linguistics for Linguistic Complexity. Colocated with CLOING 2016*. Osaka, Japan: 2016
15. Hassan Sajjad, Francisco Guzmàn, Nadir Durrani, Ahmed Abdelali, **Houda Bouamor**, Irina Temnikova and Stephan Vogel. "Eyes Don't Lie: Predicting Machine Translation Quality

- Using Eye Movement". *NAACL HLT 2016. The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, CA, USA: 2016.
16. Wajdi Zaghoulani, **Houda Bouamor**, Abdelati Hawwari, Mona Diab, Ossama Obeid, Mahmoud Ghoneim, Sawsan Alqahtani and Kemal Oflazer. *Large Scale Arabic Diacritized Corpus: Guidelines and Framework*. *LREC 2016*. Portoroz, Slovenia: 2016.
  17. Salam Khalifa, **Houda Bouamor** and Nizar Habash. *DALILA: The Dialectal Arabic Linguistic Learning Assistant*. *LREC 2016*. Portoroz, Slovenia: 2016
  18. Ossama Obeid, **Houda Bouamor**, Wajdi Zaghoulani, Mahmoud Ghoneim, Abdelati Hawwari, Mona Diab and Kemal Oflazer. *MANDIAC: A Web-based Annotation System For Manual Arabic Diacritization*. *OSACT2 2016*. Portoroz, Slovenia: 2016
  19. Alla Rozovskaya, **Houda Bouamor**, Nizar Habash, Wajdi Zaghoulani, Ossama Obeid, Behrang Mohit. *The Second QALB Shared Task on Automatic Text Correction for Arabic*. *ANLP 2015. Second Workshop on Arabic Natural Language Processing*. Beijing, China: 2015.
  20. **Houda Bouamor**, Wajdi Zaghoulani, Mona Diab, Ossama Obeid, Kemal Oflazer, Mahmoud Ghoneim, Abdelati Hawwari. *A Pilot Study on Arabic Multi-Genre Corpus Diacritization*. *ANLP 2015. Second Workshop on Arabic Natural Language Processing*. Beijing, China: 2015.
  21. **Houda Bouamor**, Hassan Sajjad, Nadir Durrani, Kemal Oflazer. *QCMUQ@QALB-2015 Shared Task: Combining Character level MT and Error-tolerant Finite-State Recognition for Arabic Spelling Correction*. *ANLP 2015. Second Workshop on Arabic Natural Language Processing*. Beijing, China: 2015.
  22. Fethi Bougares and **Houda Bouamor**. *UMMU@QALB-2015 Shared Task: Character and Word level SMT pipeline for Automatic Error Correction of Arabic Text*. *ANLP 2015. Second Workshop on Arabic Natural Language Processing*. Beijing, China: 2015.
  23. Wajdi Zaghoulani, Nizar Habash, **Houda Bouamor**, Alla Rozovskaya, Behrang Mohit, Kemal Oflazer, "Correction Annotation for Non-Native Arabic Texts: Guidelines and Corpus". *LAW IX 2015. The 9th Linguistic Annotation Workshop*, Denver, USA: 2015.
  24. **Houda Bouamor**, Hanan Alshikhabobakr, Behrang Mohit, Kemal Oflazer, "A Human Judgement Corpus and a Metric for Arabic MT Evaluation". *EMNLP 2014. Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar: 2014.
  25. Serena Jeblee, Weston Feely, **Houda Bouamor**, Alon Lavie, Nizar Habash and Kemal Oflazer. "Domain and Dialect Adaptation for Machine Translation into Egyptian Arabic". *ANLP14. Workshop on Arabic Natural Language Processing at EMNLP 2014*, Doha, Qatar: 2014.
  26. Serena Jeblee, **Houda Bouamor**, Wajdi Zaghoulani and Kemal Oflazer. "CMUQ at The 2014 Automatic Arabic Error Correction Shared Task". *ANLP 2014. Workshop on Arabic Natural Language Processing at EMNLP 2014*, Doha, Qatar: 2014.

27. Sabih Bin Wasi, Rukhsar Neyaz, **Houda Bouamor** and Behrang Mohit. "CMUQ@Qatar: Using Rich Lexical Features for Sentiment Analysis on Twitter". *Semeval'14. The Eighth International Workshop on Semantic Evaluation*, Dublin, Ireland: 2014.
28. Kamla Al-Mannai, Hanan Alshikhabobakr, Sabih Bin Wasi, Rukhsar Neyaz, **Houda Bouamor** and Behrang Mohit. "CMUQ-Hybrid: Sentiment Classification By Feature Engineering and Parameter Tuning". *Semeval'14. The Eighth International Workshop on Semantic Evaluation*, Dublin, Ireland: 2014.
29. **Houda Bouamor**, Nizar Habash, Kemal Oflazer, "A Multidialectal Parallel Corpus of Arabic" *LREC 2014. The 9th international conference on Language Resources and Evaluation*, Reykjavik, Iceland: 2014.
30. Ahmed Salama, **Houda Bouamor**, Behrang Mohit, Kemal Oflazer, "YouDACC: the Youtube Dialectal Arabic Commentary Corpus " *LREC 2014. The 9th international conference on Language Resources and Evaluation*, Reykjavik, Iceland: 2014.
31. **Houda Bouamor**, Behrang Mohit, Kemal Oflazer, "SuMT: A Framework of Summarization and MT" *IJCNLP 2013. The 6th International Joint Conference on Natural Language Processing*, Nagoya, Japan: 2013. (**Best Paper Award**)
32. Mahmoud Azab, **Houda Bouamor**, Behrang Mohit, Kemal Oflazer, " Dudley North visits North London: Learning When to Transliterate to Arabic" *NAACL-HLT 2013. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* , Atlanta, USA: 2013.
33. Aurélien Max, **Houda Bouamor**, Anne Vilnat, " Generalizing Sub-sentential Paraphrase Acquisition across Original Signal of Text Pairs" *EMNLP-CONLL 2012. Conference on Empirical Methods in Natural Language Processing and Natural Language Learning*, Jeju, Korea: 2012.
34. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Validation of Sub-sentential Paraphrases Acquired From Parallel Monolingual Corpora" *EACL 2012. The 13th Conference of the European Chapter of the Association for computational Linguistics*, Avignon, France: 2012.
35. **Houda Bouamor**, Aurélien Max, Gabriel Illouz, Anne Vilnat, "A contrastive review of paraphrase acquisition techniques" *LREC 2012. The 8th international conference on Language Resources and Evaluation*, Istanbul, Turkey: 2012.
36. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Une étude en 3D de la paraphrase : types de corpus, langues et techniques" *TALN 2012. 19ème Conférence sur le Traitement Automatique des Langues Naturelles*, Grenoble, France: 2012.
37. **Houda Bouamor**, Aurélien Max, Gabriel Illouz, Anne Vilnat, "Validation sur le Web de reformulations locales : application à la Wikipédia" *TALN 2012. 19ème Conférence sur le Traitement Automatique des Langues Naturelles*, Grenoble, France: 2012.
38. **Houda Bouamor**, Aurélien Max, Anne Vilnat, " Monolingual Alignment by Edit Rate Computation on Sentential Paraphrase Pairs" *ACL-HLT 2011. The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA: 2011.

39. **Houda Bouamor**, Aurélien Max, Gabriel Illouz, Anne Vilnat, “Web-based validation for contextual targeted paraphrasing” *Text-To-Text 2011. The ACL 2011 Workshop on Monolingual Text-To-Text Generation*, Portland, Oregon, USA: 2011.
40. **Houda Bouamor**, Aurélien Max, Anne Vilnat, “Combinaison d’informations pour l’alignement monolingue” *TALN 2011. 18ème Conférence sur le Traitement Automatique des Langues Naturelles*, Montpellier, France: 2011.
41. Camille Dutrey, **Houda Bouamor**, Delphine Bernhard, Aurélien Max, Anne Vilnat, “Paraphrases et modifications locales dans l’historique des révisions de Wikipédia” *TALN 2011. 18ème Conférence sur le Traitement Automatique des Langues Naturelles*, Montpellier, France: 2011.
42. **Houda Bouamor**, Aurélien Max, Anne Vilnat, “Acquisition de paraphrases sous-phrastiques depuis des paraphrases d’énoncés” *TALN 2010. 17ème Conférence sur le Traitement Automatique des Langues Naturelles, papier court*, Montréal, Canada: 2010.
43. **Houda Bouamor**, Aurélien Max, Anne Vilnat, “Comparison of paraphrase acquisition techniques on sentential paraphrases” *IceTAL 2010. 7th International Conference on Natural Language Processing*, Reykjavik, Iceland: 2010.
44. Adrian Popescu, Gregory Grefenstette, **Houda Bouamor**, “Mining a Multilingual Geographical Gazetteer from the Web” *WI 2009. The 2009 IEEE/WIC/ACM International Conference On Web Intelligence*, Milan, Italy: 2009.
45. **Houda Bouamor**, Aurélien Max, Anne Vilnat, “Amener des utilisateurs à créer et évaluer des paraphrases par le jeu” *TALN 2009. 16ème Conférence sur le Traitement Automatique des Langues Naturelles, Session de démonstrations*, Senlis, France: 2009.
46. **Houda Bouamor**, “Extraction des connaissances à partir du Web pour la recherche des images géoréférencées” *RJCRI 2009. 4èmes Rencontres Jeunes Chercheurs en Recherche d’Informations*, Presqu’îles de Giens, France: 2009.

## Thesis/Reports

**Houda Bouamor**, “Étude de la paraphrase sous-phrastique en traitement automatique des langues”, Université Paris Sud, École doctorale d’Informatique. PhD Thesis. 2012

Camille Dutrey, **Houda Bouamor**, Delphine Bernhard, Aurélien Max. “Typologie des modifications dans les révisions de Wikipédia”. LIMSI Technical Report 2011-01

**Houda Bouamor**, “Extraction de connaissances à partir du Web pour la recherche d’images géo-référencées sur Internet”, Université Paris-Est Marne-la-Vallée. Msc Thesis. 2008

---

## Authored Resources

Original resources available at: <http://nlp.qatar.cmu.edu/#resources>



- The MADAR Arabic Dialect Corpus and Lexicon
  - Arabic Dialect Twitter Corpus
  - MADARi: A Web Interface for Joint Arabic Morphological Annotation and Spelling Correction
  - Unified Guidelines and Resources for Arabic Dialect Orthography.
  - AL-BLEU dataset and software for evaluation of Arabic Machine Translation
  - A Multidialectal Parallel Corpus of Arabic
  - YouDACC: the Youtube Dialectal Arabic Comment Corpus
  - The 2014 Automatic Arabic Error Correction Shared Task QALB data
  - Arabic Named Entity Translation Lexicon
  - The Wikipedia Correction and Paraphrase Corpus (WiCoPaCo)
- 

## Research Grants

### Past grant

- "OptDiac: An Optimal Diacritization Scheme for Arabic Orthographic Representation". Mona Diab (Lead-PI), Kemal Oflazer and Houda Bouamor. Qatar National Research Fund Award. Ongoing . \$866,551 total spread over three years. Collaboration with The George Washington University.

### Current grant

- "MADAR: Multi-Arabic Dialect Applications and Resources". Nizar Habash (Lead-PI), **Houda Bouamor**, Kemal Oflazer and Owen Rambow . Qatar National Research Fund Award. Ongoing . \$866,551 total spread over three years. Collaboration with New York University Abu Dhabi and Columbia University.
- 

## Honors/Awards

- Anita Borg Institute Faculty Scholarship to attend The 2015 edition of the Grace Hopper Conference
  - Best paper award in the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013)
  - Full PhD scholarship from the Ministry of Higher Education and Research, France
-

## Service

### Conference/Workshop Program Committees

- Invited speaker at Social MT 2017, Prague, Czech Republic
- Invited speaker at CEC-TAL 2015, Tunisia
- Publication co-chair: EMNLP 2014 Arabic NLP Workshop, ACL 2015 Arabic NLP Workshop
- Machine Translation session chair: IJCNLP 2013
- Journal reviewer: TALIP 2013, TALIP 2014, Computer Speech and Language, Information Processing and Management, Natural Language Engineering
- Conference reviewer: COLING 2012, AICCSA 2013, IJCNLP 2013, EMNLP 2013, EACL 2014, CICLING 2014, ACL 2014, EMNLP 2014, Semeval 2014, ACL 2015, EMNLP 2015, LREC 16, EMNLP 2016, ACL 2017, EMNLP 2017, LREC 2018, IJCAI 2018, IJCAI 2018, IJCAI 2019, ACL 2019, EMNLP 2019, NAACL 2019
- Program committee member: IJCNLP 2013, AICCSA 2013, AICCSA 2014, ANLP 2014, OSACT2 2016, IJCAI-2019 FinNLP, ACL 2019, IJCAI 2019

### Thesis and project supervision

- (2019) Teodora Ostojic. Linguistics Masters internship  
*Topic: A linguistic analysis of Named Entities in German Financial documents*
- (2019) Ana Calatayud. NLP Masters internship  
*Topic: Automatic Named Entity Recognition in German Financial documents using deep learning.*
- (2019) Sonia Ben Berghouth and Oussama Bouftas. Finance Internship  
*Topic: Annotation of financial prospectuses*
- (2018) Abderrahim Ait-Azzi. Masters Research Project  
*Topic: Neural Text Generation from Structured Data with application to Finance*
- (2016) Maher Hasan Khan, Zehin Khairullah, Ameera Tag Junior Project with Selma Limam Mansar and Shadi Aoun  
*Topic: GE Power: Data analysis and Modeling for Power Resource Optimization in GE*
- (2016) Naassih Gopee thesis with Kemal Oflazer, Bhiksha Raj and William Cohen  
*Topic: Applying Recurrent Neural Network on Arabic Named Entity Recognition*
- (2016) Alaa Khader thesis with Kemal Ofalzer and Francisco Guzmàn  
*Topic: Computer Assisted Learning for Arabic Speaking ESL Students*
- (2015) Sabih Bin Wasi thesis with Kemal Oflazer and Alexander Cheek  
*Topic: Machine Learning for Twitter Sentiment Analysis*

- (2011) Engineer Internship of Jingjing Li with Aurélien Max  
*Topic: Development of a Web-based interface for writing assistance*
- (2010) Msc Internship of Camille Dutrey with Aurélien Max and Delphine Bernhard  
*Topic: Identifying local modifications and paraphrases in Wikipedia's revision history*
- (2010) Engineering Internship of Léton Attanon with Aurélien Max  
*Topic: Development of a Web-based Game for sub-sentential paraphrase acquisition*

### **Workshop Organization**

- Co-organizer of the FinSBD-2019 Shared Task on Sentence Boundary Detection in PDF Noisy Text in the Financial Domain
- Co-organizer of the MADAR Shared Task on Arabic Fine-Grained Dialect Identification 2019
- Co-organizer of the Arabic Natural Language Processing Workshop 2019
- Co-organizer of the Arabic Natural Language Processing Workshop 2017
- Co-organizer of the Workshop on Semitic Machine Translation 2016
- Co-organizer of the First Workshop on Arabic Dialect Technologies 2016
- Publication co-chair of the Arabic Natural Language Processing Workshop 2015
- Co-chair of the Second Shared Task on Automatic Arabic Error Correction 2015
- Publication co-chair of the Arabic Natural Language Processing Workshop 2014

### **University Services**

- Member of the Institutional Review Board (IRB), Carnegie Mellon University in Qatar
- Member of the Events and Outreach Committee, Carnegie Mellon University in Qatar
- Chaperone of the Silicon Valley Student Innovation Trip, Spring 2016
- Chaperone of the Vietnam Service Learning Trip - May 2017

---

### **Technical Skills**

- Machine learning: Algorithms, techniques and tools for supervised, unsupervised and deep learning (Regression, Decision Trees, Bayesian Models, SVM, KNN, CNN, LSTM, etc.)
- Machine learning toolkits: Sickit-learn, tensorflow, theano, keras
- Programming languages: Python, Java, C, Perl, PHP

- Python data analysis and visualization toolkits: numpy, scipy, pandas, matplotlib, Seaborn
- Version control systems: Git, SVN
- DBMS: Oracle, MySQL
- Information System Analysis and Design: UML, Merise
- Languages: English, French and Arabic

## References

Available upon request